AI-generated feedback on writing: insights into efficacy and ENL student preference

Abstract

The question of how generative AI tools, such as large language models and chatbots, can be leveraged ethically and effectively in education is ongoing. Given the critical role that writing plays in learning and assessment within educational institutions, it is of growing importance for educators to make thoughtful and informed decisions as to how and in what capacity generative AI tools should be leveraged to assist in the development of students' writing skills. This paper reports on two longitudinal studies. Study 1 examined learning outcomes of 48 university English as a new language (ENL) learners in a six-week long repeated measures quasi experimental design where the experimental group received writing feedback generated from ChatGPT (GPT-4) and the control group received feedback from their human tutor. Study 2 analyzed the perceptions of a different group of 43 ENLs who received feedback from both ChatGPT and their tutor. Results of study 1 showed no difference in learning outcomes between the two groups. Study 2 results revealed a near even split in preference for AI-generated or human-generated feedback, with clear advantages to both forms of feedback apparent from the data. The main implication of these studies is that the use of AI-generated feedback can likely be incorporated into ENL essay evaluation without affecting learning outcomes, although we recommend a blended approach that utilizes the strengths of both forms of feedback. The main contribution of this paper is in addressing generative AI as an automatic essay evaluator while incorporating learner perspectives.